

*Jakob Voß*

## **Vom Social Tagging zum Semantic Tagging**

### **Zusammenfassung**

Social Tagging als freie Verschlagwortung durch Nutzer im Web wird immer häufiger mit der Idee des Semantic Web in Zusammenhang gebracht. Wie beide Konzepte in der Praxis konkret zusammenkommen sollen, bleibt jedoch meist unklar. Dieser Artikel soll hier Aufklärung leisten, indem die Kombination von Social Tagging und Semantic Web in Form von Semantic Tagging mit dem Simple Knowledge Organisation System dargestellt und auf die konkreten Möglichkeiten, Vorteile und offenen Fragen der Semantischen Indexierung eingegangen wird.

## **1 Grundlagen**

Im Zusammenhang mit dem Web-2.0-Hype wurde Social Tagging in den letzten Jahren als überlegene Alternative zu traditionellen Erschließungssystemen propagiert. Obgleich die freie Verschlagwortung durch Nutzer im Web ein großes Potenzial und Mehrwert gegenüber der reinen Volltextsuche bietet, wird langsam klar, dass terminologische Kontrolle bei der Organisation und dem Retrieval von Informationen und Dokumenten durchaus hilfreich sein kann. Aus diesem Grund und wegen Schwierigkeiten beim Zustandekommen einer kritischen Masse findet zur Zeit eine Annäherung an die Idee des Semantic Web statt. Da unter der Bezeichnung ‚Semantic Tagging‘ als Kombination von ‚Semantic Web‘ und ‚Social Tagging‘ gewissermaßen zwei Buzzwords zusammenkommen, um ein neues zu bilden, ist es zunächst notwendig, einige Begriffe zu klären. Im Folgenden werden deshalb die Grundlagen von Social Tagging, Folksonomies, Semantic Web und dem Simple Knowledge Organisation System vorgestellt, um anschließend auf die konkrete Umsetzung von Semantic Tagging einzugehen.

### **1.1 Social Tagging und Folksonomies**

Social Tagging (auf Deutsch „Gemeinschaftliches Indexieren“) ist eine Form der manuellen Verschlagwortung, die im Web ab etwa 2004 durch verschiedene Diens-

te populär wurde. Die erste große Tagging-Anwendung [del.icio.us](http://del.icio.us)<sup>1</sup> ermöglichte es, öffentlich Bookmarks abzulegen und dabei mit Schlagworten (Tags) zu versehen. Dabei ist sichtbar, wann, von wem und mit welchen Tags eine Seite aufgenommen wurde und welche weiteren Seiten mit einem bestimmten Tag abgespeichert wurden. Einen zusätzlichen Mehrwert bieten Formen der Navigation und Visualisierung auf Basis der vergebenen Tags wie Browsing-Zugänge und Tag-Clouds (durch ihre Schriftgröße nach Häufigkeit gewichtete Auflistungen von Tags). Feedback-Mechanismen führen dazu, dass mit steigender Anzahl von Nutzern und Ressourcen sich ein gemeinsames Vokabular herausbildet; Thomas Vander Wal prägte dafür den Begriff ‚Folksonomy‘. Die Popularität von Anwendungen mit Tagging-Funktionalität geht einher mit einer Fülle von Forschungsliteratur zum Thema und verschiedenen Spielarten der gemeinschaftlichen Indexierung.<sup>2</sup> Konkrete Anknüpfungen an informationswissenschaftliche Ergebnisse des letzten Jahrhunderts bilden dabei die Ausnahme, obgleich Tagging prinzipiell eine Form der manuellen Indexierung ist.<sup>3</sup> Neu ist, dass die Indexierung nicht auf Experten beschränkt ist und im Web stattfindet. Die Tagging-Ereignisse aller Beteiligten werden direkt sichtbar miteinander verknüpft und können so gegenseitig genutzt werden. Zwei für das semantische Tagging relevante Dimensionen von Tagging-Systemen seien hier kurz vorgestellt: Hinsichtlich der Art der Aggregation kann ein Tag zu einer Ressource entweder nur einmal (Set-Model) oder mehrfach (Bag-Model) vergeben werden. Hinsichtlich der terminologischen Kontrolle können auf der einen Seite beliebige Tags vergeben werden (nicht-kontrolliert), während auf der anderen Seite nur aus einer Menge möglicher Tags ausgewählt wird (kontrolliert). Die Meinungen darüber, ob Systeme bei denen sich Nutzer über Tags und Tag-Zuordnungen einigen müssen (bspw. beim Kategorien-System der Wikipedia)<sup>4</sup> noch als Tagging bezeichnet werden können, gehen auseinander. Es ist davon auszugehen, dass bei Tagging-Systemen Mischformen zunehmen werden und dass letztendlich die einfache Benutzbarkeit und Verfügbarkeit über den Erfolg eines Systems entscheidet.

## 1.2 Semantic Web

Das Semantic Web ist eine Erweiterung des WWW zur einheitlichen Verknüpfung und Bereitstellung von Informationen und Aussagen über Informationen (Metadaten). Dazu werden alle Objekte (Ressourcen) mit Identifikatoren in Form von Uniform Resource Identifiern (URI) versehen, über die sie miteinander verknüpft

---

1 <http://del.icio.us> wurde 2003 gegründet und nach zwei Jahren von Yahoo gekauft.

2 Für eine Typologie verschiedener Tagging-Systeme siehe Voß, 2007a.

3 Der auf Calvin Mooers, den Begründer des Information Retrieval, zurückgehende Begriff ‚Descriptor‘ entspricht im Wesentlichen dem heutigen ‚Tag‘.

4 Das Kategorien-System der Wikipedia weist trotz einiger Besonderheiten strukturelle Gemeinsamkeiten mit herkömmlichen Erschließungssystemen auf, siehe Voß, 2006.

werden können. Im Gegensatz zu einfachen Hyperlinks sind diese Verknüpfungen typisiert; die Verknüpfungstypen (Properties) werden ebenfalls durch URIs identifiziert. Die elementaren Bestandteile des Semantic Web sind diese typisierten Verweise zwischen zwei Ressourcen oder zwischen einer Ressource und einem Zeichenkettenwert. Ressource, Property und Verweisziel bzw. -wert bilden jeweils ein ‚Tripel‘ in Form einer Aussage aus Subjekt, Prädikat und Objekt. Ein von mehreren Tripeln aufgespannter Graph bildet ein Semantisches Netz bzw. im Falle aller über das WWW verfügbarer Tripeln das Semantic Web. Zur Kodierung von Tripeln dient die formale Sprache des Resource Description Framework (RDF). RDF wird mit RDF Schema (RDFS) und der Web Ontology Language (OWL) um Regeln zur Konsistenzprüfung und zur Ableitung von neuen Informationen aus vorhandenen Tripeln (Inferenz) erweitert. Computerprogrammen, die RDF-kodierte Informationen verarbeiten und darauf Schlussfolgerungsregeln anwenden können, wird mitunter zugeschrieben, dass sie die Informationen „verstehen“; diese aus der Künstlichen Intelligenz stammende Interpretation des Semantic Web ist jedoch umstritten. Zur Angabe von RDF-Fragmenten wird im Folgenden statt der verbreiteten Syntax RDF/XML die ‚Notation 3‘ (N3) verwendet, die 1998 von Tim-Berners Lee vorgeschlagen wurde. Einzelne Aussagen werden in N3 als ‚Subjekt Prädikat Objekt‘ geschrieben und enden mit einem Punkt. Aufeinander folgende Aussagen gleichen Subjekts werden stattdessen mit Semikolon getrennt und das Subjekt wird nicht wiederholt. Zur Abkürzung von URIs bietet N3 das Konzept von Namensräumen, die einem Identifikator durch Doppelpunkt getrennt vorangestellt werden. In den Beispielen dieses Artikels steht ‚skos:‘ für den Namensraum des SKOS-Vokabulars und ‚rdfs:‘ für den Namensraum von RDF Schema sowie ‚m:‘ für den Namensraum eines fiktiven Vokabulars oder Tagging-Systems und ‚#‘ für den lokalen Namensraum der getaggten Objekte. Zu den erweiterten Möglichkeiten von N3 gehören eckige Klammern („[ ]“) für ‚blank nodes‘, das sind Ressourcen mit unbekannter oder irrelevanter URI, geschweifte Klammern („{ }“) für Aussagen über Aussagen (Reification), und das Wort ‚a‘ als Abkürzung für die Property `rdfs:type`. Die letzte Zeile aus Abb. 1 lautet in N3:

```
#FotoParkbank skos:subject m:c .  
m:c skos:prefLabel „Sitzbank“ ; skos:altLabel „Bank“ .
```

### 1.3 Das Simple Knowledge Organisation System (SKOS)

Das Simple Knowledge Organisation System ist ein RDF-Vokabular (d. h. eine Menge von Ressourcen, Properties und dazu gehörenden Regeln in RDFS/OWL) zur Kodierung kontrollierter Vokabulare wie Thesauri, Klassifikationen, Taxonomien, Schlagwortdateien etc. Derzeit wird SKOS in einem offenen Standardisierungsprozess<sup>5</sup> überarbeitet, um noch 2008 als offizielle W3C Recommendation

---

5 Siehe <http://www.w3.org/2004/02/skos/> und die dort angegebenen Mailinglisten.

verabschiedet zu werden. SKOS definiert die RDF-Klasse `skos:Concept` zur Auszeichnung von Ressourcen, die zur Indexierung eingesetzt werden können (Konzepte) und die Property `skos:subject` zur Verknüpfung einer Ressource mit einem Konzept. Abbildung 1 zeigt ein semantisches Netz, in dem zwei Objekte (`#FotoSparkasse` und `#FotoParkbank`) mit jeweils einem Konzept (`m:b` bzw. `m:c`) mittels SKOS indexiert sind. Konzepte können untereinander mit `skos:broader` und `skos:narrower` hierarchisch<sup>6</sup> verknüpft und mit `skos:inScheme` einem Vokabular (`skos:ConceptScheme`) zugeordnet werden (in Abb. 1 ausgelassen). Die Zuordnung von Zeichenketten zu Konzepten für deren Identifizierung und Anzeige geschieht mit `skos:prefLabel` für Vorzugsbenennung und `skos:altLabel` für alternative Benennungen. Vorzugsbenennungen müssen pro Sprache und Vokabular eindeutig sein. Die Trennung von Konzepten und ihren Benennungen durch terminologische Kontrolle vermeidet das beim Tagging übliche Problem von Homonymen und Synonymen. So bezieht sich die homonyme Zeichenkette „Bank“ in Abbildung 1 in erster Linie auf das Konzept `m:a`, den Oberbegriff der „Sparkasse“ (`m:b`). Gleichzeitig ist „Bank“ eine alternative Benennung der „Sitzbank“ (`m:c`).

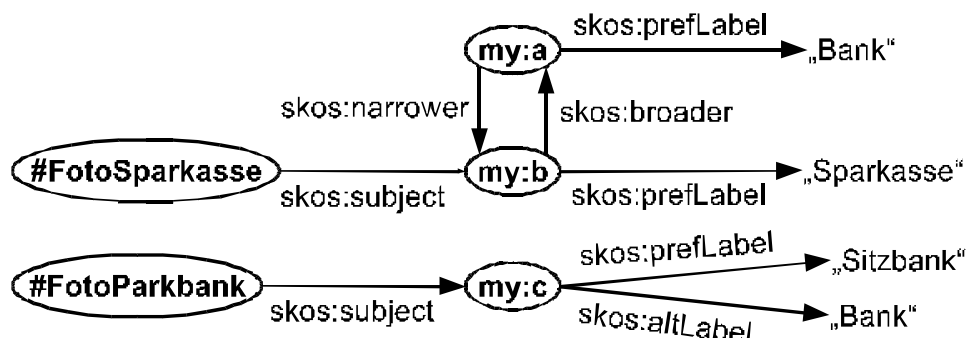


Abb. 1: Beispiel der Indexierung und Kodierung eines Vokabulars mit SKOS

Der grundlegende Vorteil der Trennung von Begriffen und ihren Benennungen mit SKOS besteht darin, dass die zur Verschlagwortung verwendeten Begriffe – im Gegensatz zu einfachen Tags – eine eindeutige URI besitzen und somit ins Semantic Web eingebunden werden können. Die Formalisierung von Informationen aus Tagging-System mit SKOS wird im folgenden Kapitel erläutert.

6 Die hierarchischen Relationen in SKOS sind übrigens nicht zwangsläufig transitiv, d. h. es findet nicht automatisch eine Anfrage-Expansion auf alle Unterbegriffe statt.

## 2 Semantisches Tagging

Um Missverständnisse zu vermeiden, sei zunächst darauf hingewiesen, dass unter ‚Semantic Tagging‘ in der Linguistik die Auszeichnung von Namen und syntaktischen Strukturen in einem Text verstanden wird – die dort vergebenen Tags entsprechen strukturell den bekannten Tags in Auszeichnungssprachen wie HTML und XML. Diese Form des Tagging mag unter Umständen als Grundlage für Tagging als Verschlagwortung herangezogen werden,<sup>7</sup> soll hier aber nicht weiter behandelt werden. Ebenfalls kann nicht tiefer darauf eingegangen werden, was genau ‚Semantik‘ ist. Angemerkt sei nur, dass in der (Sprach-)Philosophie grundsätzlich unterschieden wird zwischen Semantik als Relation von Zeichen, Wörtern, Sätzen usw. untereinander und Semantik als Relation zwischen Zeichen und der Welt.<sup>8</sup> Die Semantik des Tagging besteht in erster Linie lediglich darin, dass ein bestimmtes Objekt mit einer bestimmten Zeichenkette (dem Tag) versehen wurde und innerhalb eines Tagging-Systems mit diesem Tag wiedergefunden werden kann und dass das Tag als Markierung am Objekt erscheint.

Der Bezug des Tagging-Vorgangs zur Welt, das heißt seine praktische Relevanz und Intention kann nur außerhalb des Tagging-Systems geklärt werden, und zwar unter Bezugnahme auf die taggende Person. Da davon ausgegangen werden kann, dass sich eine Person beim Taggen nicht auf die Zeichenketten als solche, sondern auf Inhalte oder Eigenschaften der getaggten Objekte bezieht, muss unterschieden werden zwischen einem Tag und der Bedeutung, die die Person dem Tag im Moment des Tagging-Vorgangs zuschreibt.<sup>9</sup> Auf diese Bedeutung (das „Konzept“) bezieht sich die eigentliche Semantik beim Semantic Tagging und in diesem Sinne ist Tagging – abgesehen von der sinnlosen Vergabe von zufälligen Tags – immer semantisch. Wie kann nun diese Semantik mit RDF explizit gemacht werden?

Im einfachsten und allgemeinen Fall – bei der so genannten freien Verschlagwortung – muss davon ausgegangen werden, dass die Person ein Tag (zum Beispiel das homonyme Tag „Bank“) in verschiedenen Bedeutungen verwendet. Somit kann lediglich ausgedrückt werden, dass ein Objekt mit einem unbekanntem Konzept erschlossen wurde, welches die Bezeichnung „Bank“ hat:

```
#object skos:index [ rdfs:label "Bank" ] .
```

Im zweiten Fall wird davon ausgegangen, dass die Person ihre Tags konsistent verwendet und mit „Bank“ in etwa immer das gleiche meint. Es ist also bekannt,

---

7 Ein Beispiel dafür ist die Web-API Calais von Reuters (<http://www.opencalais.com/>).

8 Die Unterscheidung geht auf Gottlob Frege zurück, der bei einer Aussage ‚Sinn‘ (Art und Weise der Aussage) und ‚Bedeutung‘ (Bezug der Aussage zur Welt) unterscheidet.

9 Die Bedeutung muss sich übrigens nicht notwendig auf eine inhärente Eigenschaft der getaggte Ressource beziehen, sondern kann beispielsweise auch das Verhältnis der Person zur Ressource ausdrücken. Zu den Funktionen von Tags siehe Golder et al., S. 5.

dass ein zwar unbekanntes, aber für die betreffende Person eindeutig benennbares Konzept verwendet wurde:

```
#object skos:index [ skos:prefLabel "Bank"; skos:inScheme m:vok ] .
```

Die Resource `m:vok` bezieht sich hier auf das persönliche Tagging-Vokabular der Person. Es ist übrigens auch möglich, dass eine Person mehrere Vokabulare verwendet oder dass mehrere Personen ein gemeinsames Vokabular teilen.

Im dritten Fall lassen sich über die konsistente Verwendung eines Tags weitere Angaben über das durch ein Tag identifizierbare Konzept machen. Dem Konzept lässt sich also eine URI zugewiesen, mittels der weitere Aussagen möglich sind:

```
#resource skos:index m:a .
```

```
m:a skos:prefLabel "Bank" ; skos:inScheme m:vok ; ...
```

Die konsistente Verwendung von Tags, die die Grundlage für ein kontrolliertes Vokabular und semantisches Tagging im engeren Sinne bildet, kann auf verschiedene Weise zustande kommen. Auf der einen Seite kann eine Person Tags in einheitlicher Bedeutung verwenden und so die terminologische Kontrolle ihrer Verschlagwortung sicherstellen. Auf der anderen Seite werden vorhandene Vokabulare wie Klassifikationen und Thesauri zum Taggen verwendet, wobei die festgelegte Bedeutung der Deskriptoren weitgehend übernommen wird.

In der Mitte befinden sich die so genannten Folksonomies. Der Begriff wird Thomas Vander Wal zugeschrieben, der damit den offenen Charakter von Social-Tagging-Systemen beschrieb, an denen sich jeder beteiligen kann. Wie Clay Shirky darlegt, stoßen traditionelle Erschließungssysteme vor allem bei großen, heterogenen Gruppen von Objekten und Benutzern an ihre Grenzen. Anstatt sich nun zunächst über die genaue Bedeutung von Tags zu einigen, wird bei einer Folksonomy davon ausgegangen, dass die Verwendung eines Tags je nach Benutzer unterschiedlich ist, sich aber mit der Zeit und über die Masse durch Selbstorganisation eine einheitliche Bedeutung herauskristallisiert. Maßgeblich ist dabei die Funktion von Feedback-Mechanismen, anhand derer die Benutzer auf die gegenseitige Nutzung von Tags reagieren können. Ab wann von einer einheitlichen Verwendung ausgegangen werden kann, ist allerdings für den Allgemeinfall schwer vorherzusagen. Zwar lassen sich in einer Folksonomy Tags spezifische URIs zuweisen,<sup>10</sup> die Bedeutung der damit identifizierbaren Konzepte bleibt jedoch in der Regel eher vage. Das Problem, in großen Mengen heterogener Tagging-Ereignisse u. A. mit statistischen, probabilistischen und linguistischen Methoden einheitliche Verwendungsmuster zu erkennen und diese auf konsistente Konzepte abzubilden, dürfte genügend Aufgaben für weitere Forschungen zum Thema Social und Semantic Tagging offen lassen.

---

10 bspw. bei Technorati, siehe <http://support.technorati.com/support/siteguide/tags>.

### 3 Formen der Umsetzung

Grundsätzlich lassen sich vier verschiedene Formen angegeben werden, mit denen Social Tagging durch eine Kodierung der Beziehungen zwischen Tags und Konzepten zum Semantischen Tagging erweitert wird. Die Formen schließen sich nicht gegenseitig aus und es existieren einige Unterformen.

#### 3.1 Kontrollierte Verschlagwortung

Die kontrollierte Verschlagwortung und ihre Kodierung in SKOS wurde bereits im vorigen Kapitel vorgestellt. Wesentlich ist dabei nicht (wie oft angenommen), eine Beschränkung der Einführung neuer Tags bspw. durch Experten, sondern die konsistente Bedeutung von Tags. Semantisches Tagging ist möglich zum einen mit dem konsistente Teil der Tags eines Nutzers und zum Anderen durch die Verwendung bereits existierender Tags mit vorgegebenen Bedeutungen. Inzwischen gibt es mehrere Tagging-Dienste, bei denen auf vorhandene Vokabulare zurückgegriffen werden kann (siehe Good, 2007; Hayman, 2007) und Dienste die die Verwaltung von eigenen Tags erleichtern, um so die Konsistenz zu verbessern.<sup>11</sup> Je besser Benutzer ihre Tags auf einfache Weise umbenennen, annotieren, gruppieren und miteinander verknüpfen können, desto mehr nähern sich Tagging-Systeme der Verschlagwortung durch bzw. Verwaltung von Thesauri und anderen kontrollierten Vokabularen an. Anwendungen und Experimente zur Kombination von Tagging mit Formen kontrollierter Verschlagwortung finden sich unter anderem bei Fountopoulos (2007), Huynh et al. (2005) und Quintarelli et al. (2007).

#### 3.2 Facettierte Erschließung

Das Prinzip der Facettierung in der Wissensorganisation geht auf den indischen Bibliothekswissenschaftler Shiyali Ramamrita Ranganathan zurück, der 1933 mit der Colon-Classification (CC) das erste explizit facettierte Klassifikationssystem vorlegte. Facetten sind unterschiedliche Aspekte bei der Erschließung einer Ressource, beispielsweise Ort, Zeit und Person.<sup>12</sup> Zu den Vorteilen der Facettierung von Tags gehören die Unterscheidung von Homonymen (beispielsweise der Zeitraum „Sommer“ und der Nachname „Sommer“) und der übersichtlichere Zugang durch

---

11 Ein fundierter Vergleich von Studien zur Indexierungskonsistenz mit Ergebnissen aus Tagging-Systemen steht übrigens bislang noch aus.

12 Die Facetten der Colon-Classification sind *personality* (Personalität), *matter* (Materie), *energy* (Energie), *space* (Ort) und *time* (Zeit).

Facettenbrowsing beim Retrieval.<sup>13</sup> Eine Übersicht über den Einsatz von Facetten beim Social Tagging geben Quintarelli et al. (2007). In kontrollierten Umgebungen können Facetten durch einfache Vereinbarungen zur Tag-Syntax umgesetzt werden; beispielsweise kann vereinbart werden, Facetten den Tags als Namensraum voranzustellen, so dass „Zeit:Sommer“ als „Sommer“ in der Facette „Zeit“ interpretiert wird. Interoperabilität ist jedoch nur durch eine expliziten Kodierung von Facetten in RDF möglich. Bisher enthält der SKOS-Standard dazu keine Vorgaben oder Empfehlungen. Prinzipiell bieten sich drei Möglichkeiten an, die auch in Kombination eingesetzt werden können:

### **Spezialisierung der Indexierungsbeziehung**

Für jede Facette wird als Spezialisierung von `skos:subject` ein neues RDF-Prädikat erstellt, das anschließend zur Indexierung verwendet wird:

```
m:timeSubject  rdfs:subPropertyOf skos:subject .
m:personSubject rdfs:subPropertyOf skos:subject .
...
#urlaub m:timeSubject [ a skos:Concept ; skos:prefLabel "Sommer" ] .
#portrait m:personSubject [ a skos:Concept; skos:prefLabel "Sommer" ] .
```

### **Spezialisierung der Tags**

Für jede Facette wird als Spezialisierung von `skos:Concept` eine neue RDF-Klasse erstellt, die zur Kodierung von Konzepten verwendet wird:

```
m:Time  rdfs:subClassOf skos:Concept .
m:Person rdfs:subClassOf skos:Concept .
...
#urlaub skos:subject [ a m:Time ; skos:prefLabel "Sommer" ] .
#portrait skos:subject [ a m:Person ; skos:prefLabel "Sommer" ] .
```

### **Spezialisierung des Vokabulars**

Jede Facette wird als eigenes Vokabular aufgefasst, dem Tags zugeordnet werden:

```
m:Times  a skos:ConceptScheme .
m:People a skos:ConceptScheme .
...
#urlaubsbild skos:subject [ skos:inScheme m:Times
                           ; skos:prefLabel "Sommer" ] .
#portraitfoto skos:subject [ skos:inScheme m:People
                           ; skos:prefLabel "Sommer" ] .
```

---

13 Kombiniert wird der Zugang durch Facettenbrowsing (Eingrenzung von Treffermengen durch Auswahl in verschiedenen Facetten) oft mit Drilldown (Eingrenzung von Treffermengen durch Auswahl zusätzlicher oder speziellerer Begriffe der gleichen Facette).



Da es bei der Spezialisierung von Indexierungsbeziehung und Konzepten durch ungewollte Inferenz zu Inkonsistenzen kommen kann, ist die dritte Variante vorzuziehen. Die Behandlung der Facetten als eigenständige Vokabulare kommt ohne zusätzliche Properties und Klassen aus und kann somit von jeder SKOS-Anwendung verarbeitet werden. Allerdings ist noch unklar, wie in SKOS mehrere Facetten zu einem übergeordneten Vokabular zusammengefasst werden können, ohne dass die Eindeutigkeit der Vorzugsbenennung über alle Teilvokabulare gilt.

### 3.3 Kombination und Mapping von Vokabularen

Auf die gleiche Weise wie Tags verschiedener Facetten lassen sich beim Tagging auch Tags aus unterschiedlichen Vokabularen kombinieren. Beispielsweise können einem Anwender Tags aus einem kontrollierten Vokabular angeboten werden, während es ihm gleichzeitig freisteht, zusätzlich eigene Tags zu verwenden. Neben der Kombination verschiedener Vokabulare lassen sich beim Semantischen Tagging auch Begriffe aus verschiedenen Vokabularen aufeinander abbilden. Für solche Mappings bietet SKOS eine Reihe von Relationen an, die einzelne oder Kombinationen von Deskriptoren einander zuordnen. Beispielsweise kann mit einem Mapping das persönliche Tagging-Vokabular einer Person auf ein existierendes Vokabular abgebildet werden. Auch die Versionierung von Tags, deren Bedeutung sich geändert hat, ist durch Mappings möglich (Voß, 2007b).

### 3.4 Kodierung beliebiger Daten

Mit Verbreitung von Tagging-Diensten kam recht bald die Idee auf, Tags neben der Beschreibung des Inhalts einer Ressource auch zur Angabe konkreter Eigenschaften wie Ort und Datum zu verwenden. Zur Kodierung von Ortsangaben sind beispielsweise so genannte ‚Geotags‘ verbreitet. Dazu werden neben dem Tag „geotagged“ zwei Tags der Form „geo:lat=*latitude*“ und „geo:lon=*longitude*“ vergeben wobei *latitude* und *longitude* durch Breiten- und Längenangabe in Dezimalschreibweise nach WGS84 stehen.<sup>14</sup> Es existiert also eine spezielle Syntax, mit der die Koordinaten in Form von Tags kodiert werden. Je nach vereinbarter Syntax können nicht nur Koordinaten sondern prinzipiell beliebige Arten von Daten in Tags abgelegt werden. Für diese Form des semantischen Tagging muss jedoch jedes beteiligte System auf die jeweilige Syntax angepasst sein, so dass eine allgemeine Kodierung nicht möglich ist. Gleichwohl können aus bekannten Tag-Formaten Aussagen extrahiert und in RDF kodiert werden.<sup>15</sup>

---

14 Siehe [http://blog.makezine.com/archive/2005/07/how\\_to\\_geotagg\\_1.html](http://blog.makezine.com/archive/2005/07/how_to_geotagg_1.html).

15 Aus syntaktisch korrekten Geotags können beispielsweise Angaben im *WGS84 Geo Positioning Vocabulary* (<http://www.w3.org/2003/01/geo/>) in RDF abgeleitet werden.

## 4 Zusammenfassung und Ausblick

Mit Social Tagging wird im Web eine größere Menge von Objekten erschlossen, als sie Experten allein jemals bewältigen könnten. Um die Nachteile der heterogenen Verwendung von Tags zu verringern, werden momentan verschiedene Kombinationen mit Konzepten des Semantic Web als Semantic Tagging diskutiert und ausprobiert. Wie in diesem Artikel gezeigt wurde, spielt dabei das Simple Knowledge Organisation System (SKOS) eine wesentliche Rolle. Die darin vorhandene Trennung zwischen einfachen Tags und per URI referenzierbaren Konzepten als der Bedeutung, die eine Person einem Tag zuschreibt, macht den Kern semantischen Tagging aus.

Während im Allgemeinfall die Bedeutung eines Tags unbekannt ist, lassen sich bei konsistenter Verwendung die Tags eines Nutzers oder einer Gruppe mit SKOS zu einem kontrollierten Vokabular zusammenfassen. Dieses Vokabular kann anschließend ebenso wie andere Vokabulare zum Tagging weiterer Ressourcen eingesetzt werden. Die Abbildung zwischen mehreren Vokabularen ist ebenso möglich wie deren Kombination innerhalb eines Tagging-Vorgangs und typisierte Links (beides durch facettierte Erschließung). Erweiterungen der Tag-Syntax ermöglichen die Kodierung beliebiger Daten in Tags, beispielsweise Koordinaten.

Semantisches Tagging erleichtert nicht nur eine konsistentere Erschließung sondern auch den Austausch von Tagging-Vokabularen und -Daten zwischen unterschiedlichen Nutzern und Anwendungen. Nicht zuletzt lassen sich mit einem einheitlichen Format bereits existierende Erschließungssysteme aus ihren Datensilos befreien und im gesamten Web nutzbar machen.

Einige Details der semantischen Kodierung von Tagging-Informationen konnten im Umfang dieses Artikels nicht ausführlich behandelt werden oder sind bislang noch ungeklärt. Die Kombination mehrerer Tags zu einer (ggf. sortierten) Menge von Tags, die eine Ressource beschreiben (Postkombination)<sup>16</sup> lässt sich mit den SKOS-Klassen `skos:Collection` und `skos:SortedCollection` bewerkstelligen, wobei die Indexierung und das Mapping mit kombinierten Konzepten im aktuellen Entwurf des SKOS-Standards bislang nicht vorgesehen ist.

Das Tripel<sup>17</sup> eines Tagging-Vorgangs, bei dem Person <P> Ressource <R> mit dem Tag <T> taggt, ließe sich in SKOS folgendermaßen kodieren:

---

16 Eine Erweiterung der Postkombination (auch Postkoordination) ist die Syntaktische Indexierung, bei der sich die einer Ressource zugewiesenen Deskriptoren zusätzlich aufeinander beziehen. In RDF bedeutet dies, dass statt mit einer Menge von Konzepten mit einem Semantischen Netz verknüpfter Konzepte indexiert wird.

17 Eine häufige Modellierung von Social Tagging sind Tripel aus Ressource, Benutzer und Tag, die zusammen einen tripartiten Graph aufspannen.

```
<R> skos:subject  
[ a skos:Collection ;  
  skos:member [ a skos:Concept ; rdfs:label <T> ] ;  
  dc:creator <P> ] .
```

An die kombinierten Deskriptoren (skos:Collection) können zusätzlich weitere Metadaten zu dem Tagging-Ereignis angehängt werden, bspw. der Zeitpunkt.

Unabhängig von der Kodierung von Tagging-Vorgängen mit RDF ist in der Praxis immer wieder zu klären, wann ein Nutzer konsistent taggt und wann er mehrere Bedeutungen unter einem Tag zusammenfasst.<sup>18</sup> Diese Frage lässt sich jedoch nur im Einzelfall mittels Heuristiken und Feedback-Mechanismen lösen. Ein nicht zu unterschätzender Faktor ist dabei die einfache Benutzbarkeit der Systeme: Ebenso wie beim Social Tagging wird letztendlich die Usability darüber entscheiden, ob und wo Semantic Tagging erfolgreich angenommen und eingesetzt wird.

## Literatur

- Berners-Lee, T. & Connolly, D. (2008). *Notation3 (N3): A readable RDF syntax*. W3C Team Submission. Verfügbar unter: <http://www.w3.org/TeamSubmission/n3/> [14.01.2008]
- Fountoupoulos, G. (2007). *RichTags: A Social Semantic Tagging System*. Retrieved from: <http://eprints.ecs.soton.ac.uk/15109/> [12.2007].
- Golder, S. & Huberman, B. (2005). Usage Patterns of Collaborative Tagging Systems. *Journal of Information Science*, 32(2). 198-208.
- Good, B., Kavas, E. & Wilkinson, M. (2007). Bridging the gap between social tagging and semantic annotation. *Nature Preceedings*. Verfügbar unter: [hdl:10101/npre.2007.945.1](http://hdl.handle.net/10101/npre.2007.945.1) [überprüft 11.07.08].
- Huynh, D., Mazzocchi, S. & Karger, D. (2005). Piggy Bank: Experience the Semantic Web Inside Your Web Browser. *Int. Semantic Web Conference (ISWC)*.
- Hayman, S. (2007). Folksonomies and tagging: New developments in social book-marking. *Ark Group Conference*, Juni 2007.
- Miles, A., Bechhofer, S. (2008). SKOS Reference. W3C Working Draft. Verfügbar unter: <http://www.w3.org/TR/skos-reference> [25.01.2008].
- Quintarelli, E. et al. (2007). Facetag: Integrating Bottom-up and Top-down Classification in a Social Tagging System. *ASIS&T Bulletin*, June/July 2007.
- Shirky, C. (2005). *Ontology is Overrated – Categories, Links, and Tags*. Verfügbar unter: [http://www.shirky.com/writings/ontology\\_overrated.html](http://www.shirky.com/writings/ontology_overrated.html) [überprüft 11.07.08].
- Voß, J. (2006). Collaborative thesaurus tagging the Wikipedia way. *Wikimetrics research paper*, 1. <http://arxiv.org/abs/cs/0604036>. [überprüft 11.07.08].

---

18 Auch homonyme Tags können in SKOS kodiert werden (mit Hilfe von hierarchischen Relationen). Dazu müssen jedoch die einzelnen Bedeutungen identifizierbar sein.

- Voß, J. (2007a). Tagging, Folksonomy & Co – Renaissance of Manual Indexing?. *Proceedings of the ISI*, Cologne. Retrieved from: <http://arxiv.org/abs/cs/0701072> [überprüft 11.07.08].
- Voß, J. (2007b). Encoding changing country codes for the Semantic Web with ISO 3166 and SKOS. *Proceedings of the ECDL*, Corfu. Verfügbar unter: <http://arxiv.org/abs/0801.3908> [überprüft 11.07.08].
- Vander Wal, T. (2004). *Feed On This*. Verfügbar unter: <http://www.vanderwal.net/random/entrysel.php?blog=1562> [3.10.2004].